REMARKS

In view of the following discussion, the Applicants submit that none of the claims now pending in the application is anticipated under the provisions of 35 U.S.C. § 102 or obvious under the provisions of 35 U.S.C. § 103. Thus, the Applicants believe that all of these claims are now in allowable form.

I. REJECTION OF CLAIMS 1-3, 10-13 AND 21 UNDER 35 U.S.C. § 102

Claims 1-3, 10-13 and 21 stand rejected as being anticipated by the Pickering patent (U.S. 6,496,799, hereinafter "Pickering"). In response, the Applicants have amended independent claims 1, 11 and 21, from which claims 2-3, 10 and 12-13 depend, to more clearly recite aspects of the present invention.

Pickering teaches a voice processing system that is adapted for determining the end of a user utterance. Specifically, the system receives the user utterance, performs speech recognition processing on the utterance, and then analyzes semantic and/or prosodic properties of the user utterance to ensure that the user has effectively finished speaking before taking further action (*e.g.*, interrupting, prompting or transferring the speaker). In the case where the system analyzes prosodic features of the user utterance, this analysis may be performed subsequent to or in parallel with the speech recognition processing. Thus, if the system determines that the user utterance has effectively completed, speech recognition processing ceases, and other action, such as prompting the user for further input, is taken.

The Examiner's attention is directed to the fact that Pickering fails to disclose or suggest the novel invention of providing an endpoint signal to a speech processing application for subsequent processing of an associated speech signal, as claimed in Applicants' amended independent claims 1, 11 and 21, from which claims 2-3, 10 and 12-13 depend. Specifically, Applicants' claims 1, 11 and 21 positively recite:

1. A method for processing a speech signal comprising:
extracting prosodic features from a speech signal;
modeling the prosodic features to identify at least one speech endpoint;

producing an endpoint signal corresponding to the occurrence of the at least one speech endpoint; and

providing the endpoint signal and the speech signal to a speech processing application to facilitate subsequent processing of the speech signal. (Emphasis added)

11. Apparatus for processing a speech signal comprising:

a prosodic feature extractor for extracting prosodic features from the speech signal;

a prosodic feature analyzer for modeling the prosodic features to identify at least one speech endpoint;

an endpoint signal producer that produces an endpoint signal corresponding to the occurrence of the at least one speech endpoint; and

means for providing the endpoint signal and the speech signal to a speech processing application to facilitate subsequent processing of the speech signal. (Emphasis added)

21. An electronic storage medium for storing a program that, when executed by a processor, causes a system to perform a method for processing a speech signal comprising:

extracting prosodic features from a speech signal;

modeling the prosodic features to identify at least one speech endpoint;

producing an endpoint signal corresponding to the occurrence of the at least one speech endpoint; and

providing the endpoint signal and the speech signal to a speech processing application to facilitate subsequent processing of the speech signal. (Emphasis Added)

In one embodiment, the Applicants' invention is directed to a method for applying prosody-based endpointing to a speech signal. Conventional speech processing techniques that are used to provide signals, based on spoken words or commands (*e.g.*, for controlling devices or software programs), typically are characterized by an inability or difficulty in locating suitable speech segments within the spoken input for processing. Typical endpointing techniques identify the completion of a speech segment or utterance by measuring pauses in the given speech signal. However, since spoken language is not typically produced with such explicit indicators, typical endpointing techniques may misinterpret normal fluctuations in the rhythm of speech, such as mid-sentence pauses, to indicate the completion of an utterance. The resultant translation of a spoken command may therefore be fraught with inaccuracies.

The Applicants' invention facilitates the translation of spoken input by extracting and modeling the prosodic features of an input speech signal in order to identify at least one endpoint in the input speech signal. Output is produced in the form of an endpoint signal that represents the occurrence of the identified endpoint in the input speech signal. Both the input speech signal and the generated endpoint signal are then provided to a separate speech recognition application that uses the endpoint signal to facilitate segmentation and subsequent word recognition of the input speech signal. The resultant translated speech thus more accurately reflects the spoken input.

In contrast, Pickering teaches identifying a point at which a user utterance is effectively completed in a previously or simultaneously processed speech signal in order to improve interaction of a voice processing system with a user. Thus, Pickering fails to anticipate Applicants' invention.

Specifically, Pickering teaches a method that, at best, identifies an endpoint in a speech signal either after speech recognition processing has been performed on the speech signal, or in parallel with the speech recognition processing (*See, e.g.*, Pickering at column 8, lines 44-46: "The speech recognition is followed by a test at step 560 to determine whether or not the caller [user] has effectively finished his/her input." Emphasis added.). That is, Pickering recognizes the content of the speech signal before (or no later than simultaneously with) determining whether the speech signal contains any more useful information. This facilitates interaction with the user, but does not aid the speech recognition processing itself, since that processing has already occurred. Nowhere does Pickering teach or suggest the need to provide an endpoint signal, along with the speech signal (user utterance), to a speech processing application *e.g.*, in order to facilitate subsequent speech recognition processing of the speech signal. Pickering thus fails to anticipate a method for processing an input speech signal wherein a speech endpoint signal is provided to a speech processing application to facilitate subsequent processing of the associated speech signal, as positively claimed by the Applicants in claims 1, 11 and 21. Therefore, the Applicants submit that independent claims 1, 11 and 21 fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

Dependent claims 2-3, 10 and 12-13 depend respectively from claims 1 and 11, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 2-3, 10 and 12-13 are not anticipated by the teachings of Pickering. Therefore, the Applicants submit that dependent claims 2-3, 10 and 12-13 also fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

II. REJECTION OF CLAIMS 4-5 AND 14-15 UNDER 35 U.S.C. § 103

Claims 4-5 and 14-15 stand rejected as being obvious over Pickering in view of the Sonmez et al. article (*Modeling Dynamic Prosodic Variation For Speaker Verification*, hereinafter "Sonmez"). The Applicants respectfully traverse the rejection.

Pickering has been discussed above.

Sonmez teaches a method for automatic speaker verification by capturing suprasegmental patterns that characterize an individual's speaking style in an input speech signal. Specifically, one step of this method includes filtering out noise in the speech signal (introduced by a pitch tracker and by microintonation effects) by treating pitch tracker irregularities (*e.g.*, offshoots of the onset and the end of the speech signal) and pitch halving or doubling in raw pitch contours to extract the intonation of the speaker. This is accomplished by a piecewise-linear stylization algorithm. Features that reflect statistics of the speaker's habitual pitch movements are then extracted from the piecewise-linear model. Sonmez, like Pickering, fails to teach or suggest, however, the production of a signal in accordance with the analyzed prosodic features.

The Examiner's attention is directed to the fact that Sonmez, singularly or in combination with Pickering, fails to disclose or suggest the novel invention of providing an endpoint signal to a speech processing application for subsequent processing of an associated speech signal, as claimed in Applicants' independent claims 1 and 11, from which claims 4-5 and 14-15 depend. Applicants' claims 1 and 11 have been recited above.

As discussed above, one embodiment of the Applicants' invention is directed to method for applying prosody-based endpointing to a speech signal. The Applicants'

invention facilitates the translation of spoken input by extracting and modeling prosodic features from an input speech signal in order to identify at least one endpoint in the input speech signal. An identified endpoint is represented by an endpoint signal that is output to a speech recognition application along with the input speech signal, thereby facilitating segmentation and recognition of the input speech signal.

In contrast, Pickering and Sonmez do not, individually or in combination, teach, show or suggest a method for processing an input speech signal wherein a speech endpoint signal is provided, along with the associated speech signal, to a speech processing application to facilitate subsequent processing of the speech signal, as positively claimed by the Applicants in claims 1 and 11. Therefore, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 4-5 and 14-15 depend respectively from claims 1 and 11, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 4-5 and 14-15 are not made obvious by the teachings of Pickering in view of Sonmez. Therefore, the Applicants submit that dependent claims 4-5 and 14-15 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

III. REJECTION OF CLAIMS 6 AND 16 UNDER 35 U.S.C. § 103

Claims 6 and 16 stand rejected as being obvious over Pickering in view of Sonmez and further in view of the Shriberg et al. article (*Prosody-Based Automatic Segmentation Of Speech Into Sentences And Topics*, hereinafter "Shriberg"). The Applicants respectfully traverse the rejection.

Pickering and Sonmez have been discussed above. Shriberg teaches a method for segmenting speech signals for information extraction, topic detection or browsing/playback using prosodic information. In one embodiment, pauses are located within the speech signal, and the durations of both a pause and the words before and after the pause are analyzed to determine whether the pause represents a boundary, *e.g.*, between two topics, sentences or phrases. By identifying boundaries within the

speech signal, the method can effectively sort information contained within the speech signal.

The Examiner's attention is directed to the fact that Shriberg, singularly or in combination with Pickering and Sonmez, fails to disclose or suggest the novel invention of providing an endpoint signal to a speech processing application for subsequent processing of an associated speech signal, as claimed in Applicants' independent claims 1 and 11, from which claims 6 and 16 depend. Applicants' claims 1 and 11 have been recited above.

As discussed above, the Applicants' invention includes extracting and modeling prosodic features from an input speech signal in order to identify at least one endpoint in the input speech signal. An identified endpoint is represented by an endpoint signal that is output to a speech recognition application along with the input speech signal, thereby facilitating segmentation and recognition of the input speech signal.

In contrast, none of Pickering, Sonmez or Shriberg teaches, shows or suggests a method for processing an input speech signal wherein a speech endpoint signal is provided, along with the associated speech signal, to a speech processing application to facilitate subsequent processing of the speech signal, as positively claimed by the Applicants in claims 1 and 11. Therefore, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 6 and 16 depend from claims 1 and 11, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 6 and 16 are not made obvious by the teachings of Pickering in view of Sonmez and further in view of Shriberg. Therefore, the Applicants submit that dependent claims 6 and 16 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

IV. REJECTION OF CLAIMS 7-9 AND 17-19 UNDER 35 U.S.C. § 103

Claims 7-9 and 17-19 stand rejected as being obvious over Pickering. The Applicants respectfully traverse the rejection.

Pickering has been discussed above.

As also discussed above, Pickering fails to disclose or suggest the novel invention of providing an endpoint signal to a speech processing application for subsequent processing of an associated speech signal, as claimed in Applicants' independent claims 1 and 11, from which claims 7-9 and 17-19 depend. Applicants' claims 1 and 11 have been recited above.

Pickering in view of the Official Notice thus fails to teach or make obvious a method for processing an input speech signal wherein a speech endpoint signal is provided, along with the associated speech signal, to a speech processing application to facilitate subsequent processing of the speech signal, as positively claimed by the Applicants in claims 1 and 11. Therefore, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 7-9 and 17-19 depend from claims 1 and 11, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 7-9 and 17-19 are not made obvious by the teachings of Pickering in view of Official Notice. Therefore, the Applicants submit that dependent claims 7-9 and 17-19 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

V. REJECTION OF CLAIMS 10 AND 20 UNDER 35 U.S.C. § 103

Claims 10 and 20 stand rejected as being obvious over Pickering in view of the Shin et al. article (*Speech/Non-Speech Classification Using Multiple Features For Robust Endpoint Detection*, hereinafter "Shin"). The Applicants respectfully traverse the rejection.

Pickering has been discussed above.

Shin teaches a method for recognizing speech in noisy environments. Specifically, Shin teaches the analysis of multiple features of an input speech signal to determine whether a given frame of the speech signal can be classified as speech or non-speech (*e.g.*, noise). These features include full-band energy, band energy of audible frequency range and higher frequency range, peakyness, linear predictive

coding (LPC) residual energy and noise-filtered energy. Shin does not teach, however, that an analysis of prosodic features of the input speech signal may facilitate this determination.

The Examiner's attention is also directed to the fact that, like Pickering, Shin fails to disclose or suggest the novel invention of providing an endpoint signal to a speech processing application for subsequent processing, as claimed in Applicants' independent claims 1 and 11, from which claims 10 and 20 depend. Applicants' claims 1 and 11 have been recited above.

Pickering and Shin thus fail, singularly and in combination, to teach or make obvious a method for processing an input speech signal wherein a speech endpoint signal is provided, along with the associated speech signal, to a speech processing application to facilitate subsequent processing of the speech signal, as positively claimed by the Applicants in claims 1 and 11. Therefore, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 10 and 20 depend from claims 1 and 11, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 10 and 20 are not made obvious by the teachings of Pickering in view of Shin. Therefore, the Applicants submit that dependent claims 10 and 20 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

VI. CONCLUSION

Thus, the Applicants submit that all of the presented claims now fully satisfy the requirements of 35 U.S.C. §102 and 35 U.S.C. §103. Consequently, the Applicants believe that all of these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is

requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

7/25/05
Date

Kin-Wah Tong, Reg. No. 39,400
(732) 530- 9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
Shrewsbury, New Jersey 07702